HUMAN RPS6KA6-RELATED GENE VARIANT ASSOCIATED WITH LUNG CANCERS

FIELD OF THE INVENTION

[0001] The invention relates to the nucleic acid and polypeptide sequences of a novel human RPS6KA6-related gene variant, the preparation process thereof, and the uses of the same in diagnosing cancers, in particular, T cell lymphoblastic lymphoma.

BACKGROUND OF THE INVENTION

[0002] Lymphoma is the third most common cancer among children in the world. The major types of lymphoma are Hodgkin's and non-Hodgkin's. Non-Hodgkin's lymphoma (NHL) occurs more frequently than Hodgkin's disease among children. The major histopathological categories of NHL in children are (1) Burkitt's and Burkitt's like lymphomas; (2) lymphoblastic lymphomas; (3) anaplastic large cell lymphoma; and (4) diffuse large cell lymphomas (Percy et al., 1999). In recent years, much progress has been made toward understanding the molecular and cellular biology of NHL. Many important contributions have been made by the characterization of chromosomal translocations and the identification of several key genetic factors associated with each type of NHL (Percy et al., 1999). However, the treatments of NHL still mainly depend on chemotherapy and radiotherapy. This is because the molecular mechanisms underlying the pathogenesis of NHL remain largely unclear.

[0003] Lymphoblastic lymphoma, a predominant T-cell tumor, accounts for about 30% of childhood NHL (National Cancer Institute Cancer.gov Web site, 2004). Recent studies have shown that T cell lymphoblastic lymphoma is caused by abnormal expression of several genetic factors such as BCL-6 (Hyjek et al., (2001) Blood. 97: 270-276), MSH2/Lmo-2/Tal-1 (Lowsky et al., (1997) Blood. 89: 2276-2282) and Stat5 (Kelly et al.,

EXPRESS MAIL LABEL
NO.: EV 481667420 US

(2003) J Exp Med. 198: 79-89). Stat5 has been shown to play a role in cell cycle regulation (Nieborowska-Skorska et al., (1999) J Exp Med. 189: 1229-1242; Martino et al., (2001) J Immunol. 166: 1723-1729). Therefore, future strategies for the prevention and treatment of T cell lymphoblastic lymphoma will focus on the elucidation of genetic substrates associated with cell cycle regulation. Interestingly, three members (RSK1, RSK2, and RSK3) of the ribosomal S6 kinase (RSK) family have been shown to be involved in the cell cycle regulation and may play a role in T cell (Edelmann et al., (1996) J Biol Chem. 271: 963-71; Zhao et al., (1996) J Biol Chem. 271: 29773-29779; Brennan et al., (1999) Mol Cell Biol. 19:4729-38; Suzuki et al., (2001) J Immunol. 167:3064-73). Thus, it raised a possibility that RPS6KA6 (ribosomal S6 kinase 4; also named RSK4; GenBank accession # AF184965) has a role in the development of T cell lymphoblastic lymphoma. Therefore, the discovery of gene variants of RPS6KA6 may be important targets for diagnostic markers of T cell lymphoblastic lymphoma.

SUMMARY OF THE INVENTION

[0004] The present invention provides an RPS6KA6-related gene variant (RPS6KA6V) which is negatively expressed in human T cell lymphoblastic lymphoma. The nucleotide sequence of the gene variant and the polypeptide sequence encoded thereby can be used for the diagnosis of any diseases associated with this gene variant or T cell lymphoblastic lymphoma.

[0005] The invention further provides an expression vector and host cell for expressing the variant.

[0006] The invention further provides a method for producing the variant.

[0007] The invention further provides an antibody specifically binding to the variant.

[0008] The invention also provides methods for detecting the presence of the variant in a mammal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 shows the nucleic acid sequence (SEQ ID NO: 1) and amino acid sequence (SEQ ID NO: 2) of RPS6KA6V.

[0010] FIG. 2 shows the nucleotide sequence alignment between the human RPS6KA6 gene (SEQ ID NO: 3) and its related gene variant (RPS6KA6V).

[0011] FIG. 3 shows the amino acid sequence alignment between the human RPS6KA6 protein (SEQ ID NO: 4) and its related gene variant (RPS6KA6V).

[0012] FIG. 4 shows the semi-quantitative RT-PCR analysis of RPS6KA6V in human cell lines, wherein the left and right columns are 100 bp DNA markers.

[0013] FIG. 5 shows the semi-quantitative RT-PCR analysis of RPS6KA6V in human tissue samples, wherein the left and right columns are 100 bp DNA markers.

DETAILED DESCRIPTION OF THE INVENTION

[0014] According to the present invention, all technical and scientific terms used have the same meanings as commonly understood by persons skilled in the art.

[0015] The term "antibody" used herein denotes intact molecules (a polypeptide or group of polypeptides) as well as fragments thereof, such as Fab, $R(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies are produced by specialized B cells after stimulation by an antigen. Structurally, antibody consists of four subunits

including two heavy chains and two light chains. The internal surface shape and charge distribution of the antibody binding domain is complementary to the features of an antigen. Thus, antibody can specifically act against the antigen in an immune response.

[0016] The term "base pair (bp)" used herein denotes nucleotides composed of a purine on one strand of DNA which can be hydrogen bonded to a pyrimidine on the other strand. Thymine (or uracil) and adenine residues are linked by two hydrogen bonds. Cytosine and guanine residues are linked by three hydrogen bonds.

[0017] The term "Basic Local Alignment Search Tool (BLAST; Altschul et al., (1997) Nucleic Acids Res. 25: 3389-3402)" used herein denotes programs for evaluation of homologies between a query sequence (amino or nucleic acid) and a test sequence as described by Altschul et al. (Nucleic Acids Res. 25: 3389-3402, 1997). Specific BLAST programs are described as follows:

[0018] (1) BLASTN compares a nucleotide query sequence with a nucleotide sequence database;

[0019] (2) BLASTP compares an amino acid query sequence with a protein sequence database;

[0020] (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence with a protein sequence database;

[0021] (4) TBLASTN compares a query protein sequence with a nucleotide sequence database translated in all six reading frames; and

[0022] (5) TBLASTX compares the six-frame translations of a nucleotide query sequence with the six-frame translations of a nucleotide sequence database.

[0023] The term "cDNA" used herein denotes nucleic acids that synthesized from a mRNA template using reverse transcriptase.

[0024] The term "cDNA library" used herein denotes a library composed of complementary DNAs, which are reverse-transcribed from mRNAs.

[0025] The term "complement" used herein denotes a polynucleotide sequence capable of forming base pairing with another polynucleotide sequence. For example, the sequence 5'-ATGGACTTACT-3' binds to the complementary sequence 5'- AGTAAGTCCAT-3'.

[0026] The term "deletion" used herein denotes a removal of a portion of one or more amino acid residues/nucleotides from a gene.

[0027] The term "expressed sequence tags (ESTs)" used herein denotes short (200 to 500 base pairs) nucleotide sequence derived from either 5' or 3' end of a cDNA.

[0028] The term "expression vector" used herein denotes nucleic acid constructs which contain a cloning site for introducing the DNA into vector, one or more selectable markers for selecting vectors containing the DNA, an origin of replication for replicating the vector whenever the host cell divides, a terminator sequence, a polyadenylation signal, and a suitable control sequence which can effectively express the DNA in a suitable host. The suitable control sequence may include promoter, enhancer and other regulatory sequences necessary for directing polymerases to transcribe the DNA.

[0029] The term "host cell" used herein denotes a cell, which is used to receive, maintain, and allow the reproduction of an expression vector comprising DNA. Host cells are transformed or transfected with suitable vectors constructed using recombinant DNA methods. The recombinant DNA introduced with the vector is replicated whenever the cell divides.

[0030] The term "insertion" or "addition" used herein denotes the addition of a portion of one or more amino acid residues/nucleotides to a gene.

[0031] The term "in silico" used herein denotes a process of using computational methods (e.g., BLAST) to analyze DNA sequences.

[0032] The term "polymerase chain reaction (PCR)" used herein denotes a method which increases the copy number of a nucleic acid sequence using a DNA polymerase and a set of primers (about 20bp oligonucleotides complementary to each strand of DNA) under suitable conditions (successive rounds of primer annealing, strand elongation, and dissociation).

[0033] The term "protein" or "polypeptide" used herein denotes a sequence of amino acids in a specific order that can be encoded by a gene or by a recombinant DNA. It can also be chemically synthesized.

[0034] The term "nucleic acid sequence" or "polynucleotide" used herein denotes a sequence of nucleotide (guanine, cytosine, thymine or adenine) in a specific order that can be a natural or synthesized fragment of DNA or RNA. It may be single-stranded or double-stranded.

[0035] The term "reverse transcriptase-polymerase chain reaction (RT-PCR)" used herein denotes a process which transcribes mRNA to complementary DNA strand using reverse transcriptase followed by polymerase chain reaction to amplify the specific fragment of DNA sequences.

[0036] The term "transformation" used herein denotes a process describing the uptake, incorporation, and expression of exogenous DNA by prokaryotic host cells.

[0037] The term "transfection" used herein denotes a process describing the uptake, incorporation, and expression of exogenous DNA by eukaryotic host cells.

[0038] The term "variant" used herein denotes a fragment of sequence (nucleotide or amino acid) inserted or deleted by one or more nucleotides/amino acids.

[0039] The present invention provides the polypeptides of a novel human RPS6KA6-related gene variant, as well as the nucleic acid sequences encoding the same.

[0040] According to the present invention, human RPS6KA6 cDNA sequence was used to query the human lung EST databases (a normal lung, a large cell lung cancer, a squamous cell lung cancer and a small cell lung cancer) using BLAST program to search for RPS6KA6-related gene variants. Four ESTs showing similarity to RPS6KA6 were identified. Two were from the large cell lung cancer, one was from the squamous cell lung cancer and one was from the SCLC databases. Their corresponding cDNA clones were found to be identical after sequencing and named RPS6KA6V (RPS6KA6 variant). FIG. 1 shows the nucleic acid sequence of RPS6KA6V (SEQ ID NO: 1) and the amino acid sequence encoded thereby (SEQ ID NO: 2).

[0041] The full-length of the RPS6KA6V cDNA is a 2403bp clone containing a 2094bp open reading frame (ORF) extending from 6bp to 2099bp, which corresponds to an encoded protein of 698 amino acid residues with a predicted molecular mass of 78.2 kDa. To determine the variation in sequence of RPS6KA6V cDNA clone, an alignment of RPS6KA6 nucleotide/amino acid sequence with RPS6KA6V was performed (FIGs. 2 and 3). The results indicate that one major genetic deletion was found in the aligned sequences showing that RPS6KA6V is a 141bp deletion in the sequence of RPS6KA6 from 1978-2118bp. The lack

of 141bp (corresponding to 47aa) is an in-frame deletion in the amino acid sequence of RPS6KA6 and generates a polypeptide of 698 amino acid residues of RPS6KA6V (Fig. 3).

[0042] In the present invention, a search of ESTs deposited in dbEST (Boguski et al. (1993) Nat Genet. 4: 332-3) at NCBI was performed to determine the tissue distribution of RPS6KA6V *in silico*. The result of *in silico* Northern analysis showed that one EST (GenBank accession number AA626690) was found to confirm the absence of 141bp region on RPS6KA6V nucleotide sequence. This EST was also generated from a lung carcinoma cDNA library suggesting that the absence of 141bp nucleotide fragment located between 1977-1978bp of RPS6KA6V may serve as a useful marker for diagnosing cancers associated with this gene variant. Therefore, any nucleotide fragments comprising 1977-1978bp of RPS6KA6V may be used as probes for determining the presence of RPS6KA6V under high stringency conditions. An alternative approach is that any set of primers for amplifying the fragment containing 1977-1978bp of RPS6KA6V may be used for determining the presence of the variant.

[0043] According to the present invention, the polypeptides of the human RPS6KA6V may be produced through genetic engineering techniques. In this case, they are produced by appropriate host cells, which have been transformed by DNAs that code for the polypeptides. The nucleotide sequence encoding the polypeptide containing 657-658aa of the human RPS6KA6V is inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence in a suitable host. The nucleic acid sequence is inserted into the vector in a manner that it will be expressed under appropriate conditions (e.g., in proper orientation and correct reading frame and with appropriate expression sequences, including an RNA polymerase binding sequence and a ribosomal binding sequence).

[0044] Any method that is known to those skilled in the art may be used to construct expression vectors containing sequences encoding the polypeptide of the human RPS6KA6V and appropriate transcriptional/translational control elements. These methods may include *in vitro* recombinant DNA and synthetic techniques, and *in vivo* genetic recombinants. (See, e.g., Sambrook, J. Cold Spring Harbor Press, Plainview N.Y., ch. 4, 8, and 16-17; Ausubel, R. M. et al. (1995) Current protocols in Molecular Biology, John Wiley & Sons, New York N.Y., ch. 9, 13, and 16.)

[0045] A variety of expression vector/host systems may be utilized to express the polypeptide-coding sequence. These include, but not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vector; yeast transformed with yeast expression vector; insect cell systems infected with virus (e.g., baculovirus); plant cell system transformed with viral expression vector (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV); or animal cell system infected with virus (e.g., vaccina virus, adenovirus, etc.). Preferably, the host cell is a bacterium, and most preferably, the bacterium is *E. coli*.

[0046] Alternatively, the polypeptide of the human RPS6KA6V or the fragments thereof may be synthesized by using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269: 202 to 204). Automated synthesis may be achieved by using the ABI 431A peptide synthesizer (Perkin-Elmer).

[0047] According to the present invention, the polypeptide and nucleic acid sequence of the human RPS6KA6V can be used as immunogen and template of primers/or probes, respectively.

[0048] The present invention further provides the antibodies which specifically bind one or more out-surface epitopes of the polypeptides of the human RPS6KA6V.

[0049] According to the present invention, immunization of mammals with immunogens described herein, preferably humans, rabbits, rats, mice, sheep, goats, cows, or horses, is performed following procedures well known to those skilled in the art, for the purpose of obtaining antisera containing polyclonal antibodies or hybridoma lines secreting monoclonal antibodies.

[0050] Monoclonal antibodies can be prepared by standard techniques, given the teachings contained herein. Such techniques are disclosed, for example, in U.S. Patent Nos. 4,271,145 and 4,196,265. Briefly, an animal is immunized with the immunogen. Hybridomas are prepared by fusing spleen cells from the immunized animal with myeloma cells. The fusion products are screened for those producing antibodies that bind to the immunogen. The positive hybridoma clones are isolated, and the monoclonal antibodies are recovered from those clones.

[0051] Immunization regimens for production of both polyclonal and monoclonal antibodies are well-known in the art. The immunogen may be injected by any of a number of routes, including subcutaneous, intravenous, intraperitoneal, intradermal, intramuscular, mucosal, or a combination thereof. The immunogen may be injected in soluble form, aggregate form, attached to a physical carrier, or mixed with an adjuvant, using methods and materials well-known in the art. The antisera and antibodies may be purified using column chromatography methods well known to those skilled in the art.

[0052] According to the present invention, antibody fragments which contain specific binding sites for the polypeptides may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$

fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab')_2$ fragments.

[0053] The subject invention also provides methods for diagnosing the diseases associated with the human RPS6KA6V or T cell lymphoblastic lymphoma, by the utilization of the nucleic acid sequence, the polypeptide of the human RPS6KA6V, and the antibodies against the polypeptide.

[0054] Many gene variants have been found to be associated with diseases (Stallings-Mann et al., (1996) Proc Natl Acad Sci U S A 93: 12394-9; Liu et al., (1997) Nat Genet 16:328-9; Siffert et al., (1998) Nat Genet 18: 45 to 8; Lukas et al., (2001) Cancer Res 61: 3212 to 9). Since RPS6KA6V clone was isolated from lung cancers cDNA libraries and identified its expression in lung carcinoma cDNA library using *in silico* Northern analysis, it is advisable that RPS6KA6V may serve as a marker for the diagnosis of human cancers. Thus, the expression level of RPS6KA6V relative to RPS6KA6 may be a useful indicator for screening of patients suspected of having cancers. This suggests that the index of relative expression level (mRNA or protein) may confer an increased susceptibility to cancers. Fragments of RPS6KA6 mRNA may be detected by RT-PCR approach. Polypeptides of RPS6KA6V may be determined by the binding of antibodies to these polypeptides. These approaches may be performed in accordance with conventional methods well known to persons skilled in the art.

[0055] According to the present invention, the expression of the gene variant mRNA in sample may be determined by, but not limited to, RT-PCR. Using TRIZOL reagents (Life Technology), total RNA may be isolated from patient samples. Tissue samples (e.g., biopsy samples) are powdered under liquid nitrogen before homogenization. RNA purity and integrity are assessed by absorbance at 260/280 nm and by agarose gel

electrophoresis. A set of primers can be designed to amplify the expected size of specific PCR fragments of RPS6KA6V. For example, one of the primers may be designed to have a sequence comprising the nucleotides of SEQ ID NO: 1 containing nucleotides 1974 to 1979, and the other may be designed to have a sequence complementary to the nucleotides of SEQ ID NO: 1 at any other locations. Alternatively, one of the primers may be designed to have a sequence complementary to the nucleotides of SEQ ID NO: 1 upstream of nucleotide 1977 and the other may be designed to have a sequence comprising the nucleotides of SEQ ID NO: 1 downstream of nucleotide 1978. In this case, both RPS6KA6 and RPS6KA6V will be amplified. The length of the PCR fragment from RPS6KA6V will be 141bp shorter than that from RPS6KA6. PCR fragments are analyzed on a 1% agarose gel using five microliters (10%) of the amplified products. The intensity of the signals may be determined by using the Molecular Analyst program (version 1.4.1; Bio-Rad). Thus, the index of relative expression levels for each co-amplified PCR product may be calculated based on the intensity of signals.

[0056] The RT-PCR experiment may be performed according to the manufacturer's instructions (Boehringer Mannheim). A 50μl reaction mixture containing 2μl total RNA (0.1μg/μl), 1μl each primer (20 pM), 1μl each dNTP (10 mM), 2.5 μl DTT solution (100 mM), 10 μl 5X RT-PCR buffer, 1μl enzyme mixture, and 28.5 μl sterile distilled water may be subjected to the conditions such as reverse transcription at 60°C for 30 minutes followed by 35 cycles of denaturation at 94°C for 2 minutes, annealing at 60°C for 2 minutes, and extension at 68°C for 2 minutes. The RT-PCR analysis may be repeated twice to ensure reproducibility, for a total of three independent experiments.

[0057] The expression of the gene variant can also be analyzed using Northern Blot hybridization approach. Specific fragment of the RPS6KA6V may be amplified by polymerase chain reaction (PCR) using

primer set designed for RT-PCR. The amplified PCR fragment may be labeled and serve as a probe to hybridize the membranes containing total RNAs extracted from the samples under the conditions of 55°C in a suitable hybridization solution for 3 hr. Blots may be washed twice in 2 x SSC, 0.1% SDS at room temperature for 15 minutes each, followed by two washes in 0.1 x SSC and 0.1% SDS at 65°C for 20 minutes each. After these washes, blot may be rinsed briefly in suitable washing buffer and incubated in blocking solution for 30 minutes, and then incubated in suitable antibody solution for 30 minutes. Blots may be washed in washing buffer for 30 minutes and equilibrated in suitable detection buffer before detecting the signals. Alternatively, the presence of gene variant (cDNAs or PCR) can be detected using microarray approach. The cDNAs or PCR products corresponding to the nucleotide sequences of the present invention may be immobilized on a suitable substrate such as a glass slide. Hybridization can be preformed using the labeled mRNAs extracted from samples. After hybridization, nonhybridized mRNAs are removed. The relative abundance of each labeled transcript, hybridizing to a cDNA/PCR product immobilized on the microarray, can be determined by analyzing the scanned images.

[0058] According to the present invention, the presence of the polypeptide of the gene variant in samples may be determined by, but not limited to, the immunoassay, which uses the antibody specifically binding to the polypeptide. For instance, the polypeptide in protein samples obtained from the mammal suspected of having such diseases may be determined by, but not limited to, the immunoassay wherein the antibody specifically binding to the polypeptide of the invention is brought into contact with the protein samples, and the antibody-polypeptide complex is detected. If necessary, the amount of antibody-polypeptide complex can be determined.

[0059] The polypeptides of the human RPS6KA6V may be expressed in prokaryotic cells by using suitable prokaryotic expression vectors. The

cDNA fragments of RPS6KA6V gene encoding the amino acid coding sequence may be PCR amplified using primer set with restriction enzyme digestion sites incorporated in the 5' and 3' ends, respectively. The PCR products can then be enzyme digested, purified, and inserted into the corresponding sites of prokaryotic expression vector in-frame to generate recombinant plasmids. Sequence fidelity of this recombinant DNA can be verified by sequencing. The prokaryotic recombinant plasmids may be transformed into host cells (e.g., *E. coli* BL21 (DE3)). Recombinant protein synthesis may be stimulated by the addition of 0.4 mM isopropylthiogalactoside (IPTG) for 3h. The bacterially-expressed proteins may be purified.

[0060] The polypeptide of the gene variant may be expressed in animal cells by using eukaryotic expression vectors. Cells may be maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS; Gibco BRL) at 37°C in a humidified 5% CO_2 atmosphere. Before transfection, the nucleotide sequence of each of the gene variant may be amplified with PCR primers containing restriction enzyme digestion sites and ligated into the corresponding sites of eukaryotic expression vector in-frame. Sequence fidelity of this recombinant DNA can be verified by sequencing. The cells may be plated in 12-well plates one day before transfection at a density of 5 x 10^4 cells per well. Transfections may be carried out using Lipofectamine Plus transfection reagent according to the manufacturer's instructions (Gibco BRL). Three hours following transfection, medium containing the complexes may be replaced with fresh medium. Forty-eight hours after incubation, the cells may be scraped into lysis buffer (0.1 M Tris HCl, pH 8.0, 0.1% Triton X-100) for purification of expressed proteins. After these proteins are purified, monoclonal antibodies against these purified proteins (RPS6KA6V) may be generated using hybridoma technique according to the conventional methods (de StGroth and Scheidegger, (1980) J Immunol

Methods 35:1-21; Cote et al. (1983) Proc Natl Acad Sci U S A 80: 2026-30; and Kozbor et al. (1985) J Immunol Methods 81:31-42).

[0061] According to the present invention, the presence of the polypeptides of the gene variant in samples may be determined by, but not limited to, Western blot analysis. Proteins extracted from samples may be separated by SDS-PAGE and transferred to suitable membranes such as polyvinylidene difluoride (PVDF) in transfer buffer (25 mM Tris-HCl, pH 8.3, 192 mM glycine, 20% methanol) with a Trans-Blot apparatus for 1h at 100 V (e.g., Bio-Rad). The proteins can be immunoblotted with specific antibodies. For example, membrane blotted with extracted proteins may be blocked with suitable buffers such as 3% solution of BSA or 3% solution of nonfat milk powder in TBST buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% Tween 20) and incubated with monoclonal antibody directed against the polypeptides of gene variants. Unbound antibody is removed by washing with TBST for 5 X 1 minutes. Bound antibody may be detected using commercial ECL Western blotting detecting reagents.

[0062] The following examples are provided for illustration, but not for limiting the invention.

EXAMPLES

Analysis of Human Lung EST Databases

[0063] Expressed sequence tags (ESTs) generated from the large-scale PCR-based sequencing of the 5'-end of human lung (normal, SCLC, squamous cell lung cancer and large cell lung cancer) cDNA clones were compiled and served as EST databases. Sequence comparisons against the nonredundant nucleotide and protein databases were performed using BLASTN and BLASTX programs (Altschul et al., (1997) Nucleic Acids Res. 25: 3389-3402; Gish and States, (1993) Nat Genet 3:266-272), at the National Center for Biotechnology Information (NCBI) with a significant

cutoff of $p<10^{-10}$. ESTs representing putative RPS6KA6V gene were identified during the course of EST generation.

Isolation of cDNA Clones

[0064] Four identical cDNA clones exhibiting EST sequences similar to the RPS6KA6 gene were isolated from lung cancer cDNA libraries and named RPS6KA6V. The inserts of these clones were subsequently excised *in vivo* from the λZAP Express vector using the ExAssist/XLOLR helper phage system (Stratagene). Phagemid particles were excised by coinfecting XL1-BLUE MRF' cells with ExAssist helper phage. The excised pBluescript phagemids were used to infect *E. coli* XLOLR cells, which lack the amber suppressor necessary for ExAssist phage replication. Infected XLOLR cells were selected using kanamycin resistance. Resultant colonies contained the double stranded phagemid vector with the cloned cDNA insert. A single colony was grown overnight in LB-kanamycin, and DNA was purified using a Qiagen plasmid purification kit.

Full Length Nucleotide Sequencing and Database Comparisons

[0065] Phagemid DNA was sequenced using the Epicentre#SE9101LC SequiTherm EXCEL™II DNA Sequencing Kit for 4200S-2 Global NEW IR^2 DNA sequencing system (LI-COR). Using the primer-walking approach, full-length sequence was determined. Nucleotide and protein searches were performed using BLAST against the non-redundant database of NCBI.

In Silico Tissue Distribution (Northern) Analysis

[0066] The coding sequence for each cDNA clones was searched against the dbEST sequence database (Boguski et al., (1993) Nat Genet. 4: 332-3) using the BLAST algorithm at the NCBI website. ESTs derived from each tissue were used as a source of information for transcript tissue expression analysis. Tissue distribution for each isolated cDNA clone was determined

by ESTs matching to that particular sequence variants (insertions or deletions) with a significance cutoff of $p<10^{-10}$.

Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) Analysis

[0067] Total RNA was extracted from WI-38 (fibroblast), A549 (lung adenocarcinoma), NCI-H661 (lung large cell carcinoma), NCI-H520 (lung squamous carcinoma), NCI-H209 (lung small cell carcinoma), JHH-4 (hepatoma), SUP-T1 (T-cell lymphoblastic lymphoma), Daudi (Burkitt's lymphoma), Ramos (Burkitt's lymphoma), and Raji (Burkitt's lymphoma) cell lines and from four breast cancer, two gastric ulcer, two colon cancer, two hepatoma, one Grave's disease, one colon cancer, one pancreatic carcinoma, one left neck tumor, one gastric carcinoma, two thyroid tumor, one spleen, one pancreatic abscess, one Gastric carcinoma, one adenomatous polyposis, one right neck lymph tissue, one liver cirrhosis, and two parotid gland mixed tumor biopsied samples, RNA purity and integrity were assessed by the absorbance at 260/280 nm and by agarose gel electrophoresis.

[0068] The forward and reverse primers for RPS6KA6V were 5'-GGAGCAAAGGGAGCAATGGTTG-3' (SEQ ID NO: 5) and 5'-TCTTCATCCAGTTTGGCCTAGG-3' (SEQ ID NO: 6), respectively. The expected size of the specific PCR fragment was 170 bp. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH; accession No. M33197) was used as an internal control. The forward and reverse primers for GAPDH were 5'-TGGGTGTGAACCATGAGAAG-3' (SEQ ID NO: 7) and 5'-GTGTCGCTGTTGAAGTCAGA-3' (SEQ ID NO: 8), respectively. The expected size of the PCR fragment was 472 bp. The electrophoresis results of the RPS6KA6V mRNA expression patterns in 10 cell lines and 25 biopsied samples determined by RT-PCR are shown in Figs. 4 and 5. The results showed that RPS6KA6V mRNA was consistently expressed in all cell lines and tissues investigated except in the T-cell lymphoblastic

lymphoma cell line. This suggests that RPS6KA6V can be used for diagnosing T-cell lymphoblastic lymphoma when RPS6KA6V mRNA cannot be detected.

REFERENCES

Altschul et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res, 25: 3389-3402, (1997).

Ausubel et al., Current protocols in Molecular Biology, John Wiley & Sons, New York N.Y., ch. 9, 13, and 16, (1995).

Boguski et al., dbEST--database for "expressed sequence tags". Nat Genet. 4: 332-3, (1993).

Brennan et al., p70(s6k) integrates phosphatidylinositol 3-kinase and rapamycin-regulated signals for E2F regulation in T lymphocytes. Mol Cell Biol. 19:4729-38, (1999).

Cote et al., Generation of human monoclonal antibodies reactive with cellular antigens, Proc Natl Acad Sci U S A 80: 2026-30 (1983).

de StGroth and Scheidegger, Production of monoclonal antibodies: strategy and tactics, J Immunol Methods 35:1-21, (1980).

Edelmann et al., Cell cycle regulation of p70 S6 kinase and p42/p44 mitogen-activated protein kinases in Swiss mouse 3T3 fibroblasts. J Biol Chem. 271: 963-71, (1996).

Frodin and Gammeltoft, Role and regulation of 90 kDa ribosomal S6 kinase (RSK) in signal transduction. Mol Cell Endocrinol, 151:65-77, (1999).

Gish and States, Identification of protein coding regions by database similarity search, Nat Genet, 3:266-272, (1993).

Hyjek et al., BCL-6 protein is expressed in precursor T-cell lymphoblastic lymphoma and in prenatal and postnatal thymus. Blood. 97: 270-276 (2001).

Kelly et al., Stat5 synergizes with T cell receptor/antigen stimulation in the development of lymphoblastic lymphoma. J Exp Med. 198: 79-89, (2003).

Kozbor et al., Specific immunoglobulin production and enhanced tumorigenicity following ascites growth of human hybridomas, J Immunol Methods, 81:31-42 (1985).

Liu et al., Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. Nat Genet 16:328-9, (1997).

Lowsky et al., Defects of the mismatch repair gene MSH2 are implicated in the development of murine and human lymphoblastic lymphomas and are associated with the aberrant expression of rhombotin-2 (Lmo-2) and Tal-1 (SCL). Blood. 89: 2276-2282, (1997).

Lukas et al., Alternative and aberrant messenger RNA splicing of the mdm2 oncogene in invasive breast cancer. Cancer Res 61:3212-9, (2001).

Martino et al., J Immunol. Stat5 and Sp1 regulate transcription of the cyclin D2 gene in response to IL-2. 166: 1723-9, (2001).

National Cancer Institute Cancer.gov Web site (http://www.nci.nih.gov/cancerinfo/pdq/treatment/child-nonhodgkins/healthprofessional#Section_14). Childhood Non-Hodgkin's Lymphoma (PDQ®): Treatment (Last Modified: 02/18/2004)

Nieborowska-Skorska et al., Signal transducer and activator of transcription (STAT)5 activation by BCR/ABL is dependent on intact Src homology (SH)3 and SH2 domains of BCR/ABL and is required for leukemogenesis. J Exp Med. 189: 1229-1242, (1999).

Percy CL, Smith MA, Linet M, et al.: Lymphomas and reticuloendothelial neoplasms. In: Ries LA, Smith MA, Gurney JG, et al., eds.: Cancer incidence and survival among children and adolescents: United States SEER Program 1975-1995. Bethesda, Md: National Cancer Institute, SEER Program, 1999. NIH Pub.No. 99-4649. Also available online. Last accessed

September 03, 2003, pp 35-50. Last accessed April 21, 2003.

Roberge et al., A strategy for a convergent synthesis of N-linked glycopeptides on a solid support. Science 269:202-4, (1995).

Sambrook, J. Cold Spring Harbor Press, Plainview N.Y., ch. 4, 8, and 16-17.

Siffert et al., Association of a human G-protein beta3 subunit variant with hypertension. Nat Genet, 18:45-8, (1998).

Stallings-Mann et al., Alternative splicing of exon 3 of the human growth hormone receptor is the result of an unusual genetic polymorphism. Proc Natl Acad Sci U S A 93:12394-9, (1996).

Suzuki et al., Diverse transcriptional response of CD4(+) T cells to stromal cell-derived factor (SDF)-1: cell survival promotion and priming effects of SDF-1 on CD4(+) T cells. J Immunol. 167:3064-73, (2001).

Wilson, R.K. GenBank accession # AA626690

Yntema et al., GenBank accession # AF184965

Yntema et al., A novel ribosomal S6-kinase (RSK4; RPS6KA6) is commonly deleted in patients with complex X-linked mental retardation. Genomics, 62:332-43, (1999).

Zhao et al., Regulation and interaction of pp90(rsk) isoforms with mitogen-activated protein kinases. J Biol Chem. 271: 29773-29779, (1996).

SEQUENCE LISTING

<110> Dai, Ken-Shwo

<120> HUMAN RPS6KA6-RELATED GENE VARIANT ASSOCIATED WITH LUNG CANCERS

<130> U 015086-3

<140> --
<141> 2004-03-23

<160> 8

<170> PatentIn version 3.1

<210> 1
<211> 2403
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (6)..(2099)
<223>

<400> 1
gggaa atg cta cca ttc gct cct cag gac gag ccc tgg gac cga gaa atg 50
Met Leu Pro Phe Ala Pro Gln Asp Glu Pro Trp Asp Arg Glu Met
1 5 10 15

gaa gtg ttc agc ggc ggc ggc gcg agc agc ggc gag gta aat ggt ctt 98
Glu Val Phe Ser Gly Gly Gly Ala Ser Ser Gly Glu Val Asn Gly Leu
20 25 30

aaa atg gtt gat gag cca atg gaa gag gga gaa gca gat tct tgt cat 146
Lys Met Val Asp Glu Pro Met Glu Glu Gly Glu Ala Asp Ser Cys His
35 40 45

gat gaa gga gtt gtt aaa gaa atc cct att act cat cat gtt aag gaa 194
Asp Glu Gly Val Val Lys Glu Ile Pro Ile Thr His His Val Lys Glu
50 55 60

ggc tat gag aaa gca gat cct gca cag ttt gag ttg ctc aag gtt ctt 242
Gly Tyr Glu Lys Ala Asp Pro Ala Gln Phe Glu Leu Leu Lys Val Leu
65 70 75

ggt cag ggg tca ttt gga aag gtt ttt ctt gtt aga aag aag acc ggt 290
Gly Gln Gly Ser Phe Gly Lys Val Phe Leu Val Arg Lys Lys Thr Gly
80 85 90 95

cct gat gct ggg cag ctc tat gca atg aag gtg tta aaa aaa gcc tct 338
Pro Asp Ala Gly Gln Leu Tyr Ala Met Lys Val Leu Lys Lys Ala Ser
100 105 110

tta aaa gtt cga gac aga gtt cgg aca aag atg gag agg gat ata ctg 386
Leu Lys Val Arg Asp Arg Val Arg Thr Lys Met Glu Arg Asp Ile Leu
115 120 125

gtg gaa gta aat cat cca ttt att gtc aaa ttg cac tat gcc ttt cag 434
Val Glu Val Asn His Pro Phe Ile Val Lys Leu His Tyr Ala Phe Gln
130 135 140

EXPRESS MAIL LABEL
NO.: EV 481667420 US

```
act gaa ggg aaa ctg tac tta ata ctg gat ttt ctc agg gga gga gat        482
Thr Glu Gly Lys Leu Tyr Leu Ile Leu Asp Phe Leu Arg Gly Gly Asp
    145                 150                 155

gtt ttc aca aga tta tcc aaa gag gtt ctg ttt aca gag gaa gat gtg        530
Val Phe Thr Arg Leu Ser Lys Glu Val Leu Phe Thr Glu Glu Asp Val
160                 165                 170                 175

aaa ttc tac ctc gca gaa ctg gcc ctt gct ttg gat cat ctg cac caa        578
Lys Phe Tyr Leu Ala Glu Leu Ala Leu Ala Leu Asp His Leu His Gln
                180                 185                 190

tta gga att gtt tat aga gac ctg aag cca gaa aac att ttg ctt gat        626
Leu Gly Ile Val Tyr Arg Asp Leu Lys Pro Glu Asn Ile Leu Leu Asp
            195                 200                 205

gaa ata gga cat atc aaa tta aca gat ttt gga ctc agc aag gag tca        674
Glu Ile Gly His Ile Lys Leu Thr Asp Phe Gly Leu Ser Lys Glu Ser
        210                 215                 220

gta gat caa gaa aag aag gct tac tca ttt tgt ggt aca gta gag tat        722
Val Asp Gln Glu Lys Lys Ala Tyr Ser Phe Cys Gly Thr Val Glu Tyr
    225                 230                 235

atg gct cct gaa gta gta aat agg aga ggc cat tcc cag agt gct gat        770
Met Ala Pro Glu Val Val Asn Arg Arg Gly His Ser Gln Ser Ala Asp
240                 245                 250                 255

tgg tgg tca tat ggt gtt ctt atg ttt gaa atg ctt act ggt act ctg        818
Trp Trp Ser Tyr Gly Val Leu Met Phe Glu Met Leu Thr Gly Thr Leu
                260                 265                 270

cca ttt caa ggt aaa gac aga aat gag acc atg aat atg ata tta aaa        866
Pro Phe Gln Gly Lys Asp Arg Asn Glu Thr Met Asn Met Ile Leu Lys
            275                 280                 285

gca aaa ctt gga atg cct caa ttt ctt agt gct gaa gca caa agt ctt        914
Ala Lys Leu Gly Met Pro Gln Phe Leu Ser Ala Glu Ala Gln Ser Leu
        290                 295                 300

cta agg atg tta ttc aaa agg aat cca gca aat aga ttg gga tca gaa        962
Leu Arg Met Leu Phe Lys Arg Asn Pro Ala Asn Arg Leu Gly Ser Glu
    305                 310                 315

gga gtt gaa gaa atc aaa aga cat ctg ttt ttt gca aat att gac tgg       1010
Gly Val Glu Glu Ile Lys Arg His Leu Phe Phe Ala Asn Ile Asp Trp
320                 325                 330                 335

gat aaa tta tat aaa aga gaa gtt caa cct cct ttc aaa cct gct tct       1058
Asp Lys Leu Tyr Lys Arg Glu Val Gln Pro Pro Phe Lys Pro Ala Ser
                340                 345                 350

gga aaa cca gat gat act ttt tgt ttt gat cct gaa ttt act gca aaa       1106
Gly Lys Pro Asp Asp Thr Phe Cys Phe Asp Pro Glu Phe Thr Ala Lys
            355                 360                 365

aca cct aaa gat tct ccc ggt ttg cca gcc agt gca aat gct cat cag       1154
Thr Pro Lys Asp Ser Pro Gly Leu Pro Ala Ser Ala Asn Ala His Gln
        370                 375                 380
```

ctc ttc aaa gga ttc agc ttt gtt gca act tct att gca gaa gaa tat 1202
Leu Phe Lys Gly Phe Ser Phe Val Ala Thr Ser Ile Ala Glu Glu Tyr
385 390 395

aaa atc act cct atc aca agt gca aat gta tta cca att gtt cag ata 1250
Lys Ile Thr Pro Ile Thr Ser Ala Asn Val Leu Pro Ile Val Gln Ile
400 405 410 415

aat gga aat gct gca caa ttt ggt gaa gta tat gaa ttg aag gag gat 1298
Asn Gly Asn Ala Ala Gln Phe Gly Glu Val Tyr Glu Leu Lys Glu Asp
420 425 430

att ggt gtt ggc tcc tac tct gtt tgc aag cga tgc ata cat gca act 1346
Ile Gly Val Gly Ser Tyr Ser Val Cys Lys Arg Cys Ile His Ala Thr
435 440 445

acc aac atg gaa ttt gca gtg aag atc att gac aaa agt aag cga gac 1394
Thr Asn Met Glu Phe Ala Val Lys Ile Ile Asp Lys Ser Lys Arg Asp
450 455 460

cct tca gaa gag att gaa ata ttg atg cgc tat gga caa cat ccc aac 1442
Pro Ser Glu Glu Ile Glu Ile Leu Met Arg Tyr Gly Gln His Pro Asn
465 470 475

att att act ttg aag gat gtc ttt gat gat ggt aga tat gtt tac ctt 1490
Ile Ile Thr Leu Lys Asp Val Phe Asp Asp Gly Arg Tyr Val Tyr Leu
480 485 490 495

gtt acg gat tta atg aaa gga gga gag tta ctt gac cgt att ctc aaa 1538
Val Thr Asp Leu Met Lys Gly Gly Glu Leu Leu Asp Arg Ile Leu Lys
500 505 510

caa aaa tgt ttc tcg gaa cgg gag gct agt gat ata cta tat gta ata 1586
Gln Lys Cys Phe Ser Glu Arg Glu Ala Ser Asp Ile Leu Tyr Val Ile
515 520 525

agt aag aca gtt gac tat ctt cat tgt caa gga gtt gtt cat cgt gat 1634
Ser Lys Thr Val Asp Tyr Leu His Cys Gln Gly Val Val His Arg Asp
530 535 540

ctt aaa cct agt aat att tta tac atg gat gaa tca gcc agt gca gat 1682
Leu Lys Pro Ser Asn Ile Leu Tyr Met Asp Glu Ser Ala Ser Ala Asp
545 550 555

tca atc agg ata tgt gat ttt ggg ttt gca aaa caa ctt cga gga gaa 1730
Ser Ile Arg Ile Cys Asp Phe Gly Phe Ala Lys Gln Leu Arg Gly Glu
560 565 570 575

aat gga ctt ctc tta act cca tgc tac act gca aac ttt gtt gca cct 1778
Asn Gly Leu Leu Leu Thr Pro Cys Tyr Thr Ala Asn Phe Val Ala Pro
580 585 590

gag gtt ctt atg caa cag gga tat gat gct gct tgt gat atc tgg agt 1826
Glu Val Leu Met Gln Gln Gly Tyr Asp Ala Ala Cys Asp Ile Trp Ser
595 600 605

tta gga gtc ctt ttt tac aca atg ttg gct ggc tac act cca ttt gct 1874
Leu Gly Val Leu Phe Tyr Thr Met Leu Ala Gly Tyr Thr Pro Phe Ala
610 615 620

aat ggc ccc aat gat act cct gaa gag ata ctg ctg cgt ata ggc aat 1922

Asn Gly Pro Asn Asp Thr Pro Glu Glu Ile Leu Leu Arg Ile Gly Asn
625 630 635

gga aaa ttc tct ttg agt ggt gga aac tgg gac aat att tca gac gga 1970
Gly Lys Phe Ser Leu Ser Gly Gly Asn Trp Asp Asn Ile Ser Asp Gly
640 645 650 655

gca aag gga gca atg gtt gca aca tac tct gcc ctg act cac aag acc 2018
Ala Lys Gly Ala Met Val Ala Thr Tyr Ser Ala Leu Thr His Lys Thr
660 665 670

ttt caa cca gtc cta gag cct gta gct gct tca agc tta gcc cag cga 2066
Phe Gln Pro Val Leu Glu Pro Val Ala Ala Ser Ser Leu Ala Gln Arg
675 680 685

cgg agc atg aaa aag cga aca tca act ggc ctg taagatttgt ggtgttccta 2119
Arg Ser Met Lys Lys Arg Thr Ser Thr Gly Leu
690 695

ggccaaactg gatgaagatg aaattaaatg tgtggctttt ttcctattct tatcaaaggc 2179

atcgttgtct gctaaattac ttgaatatta agtaatatta aatccccatt tttaggggaa 2239

gtgagattta aaaaaccatt cacaggtcca caatattcat actatgtgtt tgcagtagtg 2299

ttcaagtgtt tatttaagca tataattggt gtccaccagg tcctcacaac ttctctgcac 2359

acaagcttct aaaattcctt tcaaataaag ttactttaat attt 2403

<210> 2
<211> 698
<212> PRT
<213> Homo sapiens

<400> 2

Met Leu Pro Phe Ala Pro Gln Asp Glu Pro Trp Asp Arg Glu Met Glu
1 5 10 15

Val Phe Ser Gly Gly Gly Ala Ser Ser Gly Glu Val Asn Gly Leu Lys
20 25 30

Met Val Asp Glu Pro Met Glu Glu Gly Glu Ala Asp Ser Cys His Asp
35 40 45

Glu Gly Val Val Lys Glu Ile Pro Ile Thr His His Val Lys Glu Gly
50 55 60

Tyr Glu Lys Ala Asp Pro Ala Gln Phe Glu Leu Leu Lys Val Leu Gly
65 70 75 80

Gln Gly Ser Phe Gly Lys Val Phe Leu Val Arg Lys Lys Thr Gly Pro
85 90 95

Asp Ala Gly Gln Leu Tyr Ala Met Lys Val Leu Lys Lys Ala Ser Leu
100 105 110

Lys Val Arg Asp Arg Val Arg Thr Lys Met Glu Arg Asp Ile Leu Val
115 120 125

Glu Val Asn His Pro Phe Ile Val Lys Leu His Tyr Ala Phe Gln Thr
130 135 140

Glu Gly Lys Leu Tyr Leu Ile Leu Asp Phe Leu Arg Gly Gly Asp Val
145 150 155 160

Phe Thr Arg Leu Ser Lys Glu Val Leu Phe Thr Glu Glu Asp Val Lys
165 170 175

Phe Tyr Leu Ala Glu Leu Ala Leu Ala Leu Asp His Leu His Gln Leu
180 185 190

Gly Ile Val Tyr Arg Asp Leu Lys Pro Glu Asn Ile Leu Leu Asp Glu
195 200 205

Ile Gly His Ile Lys Leu Thr Asp Phe Gly Leu Ser Lys Glu Ser Val
210 215 220

Asp Gln Glu Lys Lys Ala Tyr Ser Phe Cys Gly Thr Val Glu Tyr Met
225 230 235 240

Ala Pro Glu Val Val Asn Arg Arg Gly His Ser Gln Ser Ala Asp Trp
245 250 255

Trp Ser Tyr Gly Val Leu Met Phe Glu Met Leu Thr Gly Thr Leu Pro
260 265 270

Phe Gln Gly Lys Asp Arg Asn Glu Thr Met Asn Met Ile Leu Lys Ala
275 280 285

Lys Leu Gly Met Pro Gln Phe Leu Ser Ala Glu Ala Gln Ser Leu Leu
290 295 300

Arg Met Leu Phe Lys Arg Asn Pro Ala Asn Arg Leu Gly Ser Glu Gly
305 310 315 320

Val Glu Glu Ile Lys Arg His Leu Phe Phe Ala Asn Ile Asp Trp Asp
325 330 335

Lys Leu Tyr Lys Arg Glu Val Gln Pro Pro Phe Lys Pro Ala Ser Gly

340 345 350

Lys Pro Asp Asp Thr Phe Cys Phe Asp Pro Glu Phe Thr Ala Lys Thr
355 360 365

Pro Lys Asp Ser Pro Gly Leu Pro Ala Ser Ala Asn Ala His Gln Leu
370 375 380

Phe Lys Gly Phe Ser Phe Val Ala Thr Ser Ile Ala Glu Glu Tyr Lys
385 390 395 400

Ile Thr Pro Ile Thr Ser Ala Asn Val Leu Pro Ile Val Gln Ile Asn
405 410 415

Gly Asn Ala Ala Gln Phe Gly Glu Val Tyr Glu Leu Lys Glu Asp Ile
420 425 430

Gly Val Gly Ser Tyr Ser Val Cys Lys Arg Cys Ile His Ala Thr Thr
435 440 445

Asn Met Glu Phe Ala Val Lys Ile Ile Asp Lys Ser Lys Arg Asp Pro
450 455 460

Ser Glu Glu Ile Glu Ile Leu Met Arg Tyr Gly Gln His Pro Asn Ile
465 470 475 480

Ile Thr Leu Lys Asp Val Phe Asp Asp Gly Arg Tyr Val Tyr Leu Val
485 490 495

Thr Asp Leu Met Lys Gly Gly Glu Leu Leu Asp Arg Ile Leu Lys Gln
500 505 510

Lys Cys Phe Ser Glu Arg Glu Ala Ser Asp Ile Leu Tyr Val Ile Ser
515 520 525

Lys Thr Val Asp Tyr Leu His Cys Gln Gly Val Val His Arg Asp Leu
530 535 540

Lys Pro Ser Asn Ile Leu Tyr Met Asp Glu Ser Ala Ser Ala Asp Ser
545 550 555 560

Ile Arg Ile Cys Asp Phe Gly Phe Ala Lys Gln Leu Arg Gly Glu Asn
565 570 575

Gly Leu Leu Leu Thr Pro Cys Tyr Thr Ala Asn Phe Val Ala Pro Glu
580 585 590

```
Val Leu Met Gln Gln Gly Tyr Asp Ala Ala Cys Asp Ile Trp Ser Leu
        595                 600                 605

Gly Val Leu Phe Tyr Thr Met Leu Ala Gly Tyr Thr Pro Phe Ala Asn
    610                 615                 620

Gly Pro Asn Asp Thr Pro Glu Glu Ile Leu Leu Arg Ile Gly Asn Gly
625                 630                 635                 640

Lys Phe Ser Leu Ser Gly Gly Asn Trp Asp Asn Ile Ser Asp Gly Ala
                645                 650                 655

Lys Gly Ala Met Val Ala Thr Tyr Ser Ala Leu Thr His Lys Thr Phe
            660                 665                 670

Gln Pro Val Leu Glu Pro Val Ala Ala Ser Ser Leu Ala Gln Arg Arg
        675                 680                 685

Ser Met Lys Lys Arg Thr Ser Thr Gly Leu
    690                 695


<210>  3
<211>  2235
<212>  DNA
<213>  Homo sapiens

<220>
<221>  CDS
<222>  (1)..(2235)
<223>

<400>  3
atg cta cca ttc gct cct cag gac gag ccc tgg gac cga gaa atg gaa       48
Met Leu Pro Phe Ala Pro Gln Asp Glu Pro Trp Asp Arg Glu Met Glu
1               5                   10                  15

gtg ttc agc ggc ggc ggc gcg agc agc ggc gag gta aat ggt ctt aaa       96
Val Phe Ser Gly Gly Gly Ala Ser Ser Gly Glu Val Asn Gly Leu Lys
            20                  25                  30

atg gtt gat gag cca atg gaa gag gga gaa gca gat tct tgt cat gat      144
Met Val Asp Glu Pro Met Glu Glu Gly Glu Ala Asp Ser Cys His Asp
        35                  40                  45

gaa gga gtt gtt aaa gaa atc cct att act cat cat gtt aag gaa ggc      192
Glu Gly Val Val Lys Glu Ile Pro Ile Thr His His Val Lys Glu Gly
    50                  55                  60

tat gag aaa gca gat cct gca cag ttt gag ttg ctc aag gtt ctt ggt      240
Tyr Glu Lys Ala Asp Pro Ala Gln Phe Glu Leu Leu Lys Val Leu Gly
65                  70                  75                  80
```

cag ggg tca ttt gga aag gtt ttt ctt gtt aga aag aag acc ggt cct 288
Gln Gly Ser Phe Gly Lys Val Phe Leu Val Arg Lys Lys Thr Gly Pro
85 90 95

gat gct ggg cag ctc tat gca atg aag gtg tta aaa aaa gcc tct tta 336
Asp Ala Gly Gln Leu Tyr Ala Met Lys Val Leu Lys Lys Ala Ser Leu
100 105 110

aaa gtt cga gac aga gtt cgg aca aag atg gag agg gat ata ctg gtg 384
Lys Val Arg Asp Arg Val Arg Thr Lys Met Glu Arg Asp Ile Leu Val
115 120 125

gaa gta aat cat cca ttt att gtc aaa ttg cac tat gcc ttt cag act 432
Glu Val Asn His Pro Phe Ile Val Lys Leu His Tyr Ala Phe Gln Thr
130 135 140

gaa ggg aaa ctg tac tta ata ctg gat ttt ctc agg gga gga gat gtt 480
Glu Gly Lys Leu Tyr Leu Ile Leu Asp Phe Leu Arg Gly Gly Asp Val
145 150 155 160

ttc aca aga tta tcc aaa gag gtt ctg ttt aca gag gaa gat gtg aaa 528
Phe Thr Arg Leu Ser Lys Glu Val Leu Phe Thr Glu Glu Asp Val Lys
165 170 175

ttc tac ctc gca gaa ctg gcc ctt gct ttg gat cat ctg cac caa tta 576
Phe Tyr Leu Ala Glu Leu Ala Leu Ala Leu Asp His Leu His Gln Leu
180 185 190

gga att gtt tat aga gac ctg aag cca gaa aac att ttg ctt gat gaa 624
Gly Ile Val Tyr Arg Asp Leu Lys Pro Glu Asn Ile Leu Leu Asp Glu
195 200 205

ata gga cat atc aaa tta aca gat ttt gga ctc agc aag gag tca gta 672
Ile Gly His Ile Lys Leu Thr Asp Phe Gly Leu Ser Lys Glu Ser Val
210 215 220

gat caa gaa aag aag gct tac tca ttt tgt ggt aca gta gag tat atg 720
Asp Gln Glu Lys Lys Ala Tyr Ser Phe Cys Gly Thr Val Glu Tyr Met
225 230 235 240

gct cct gaa gta gta aat agg aga ggc cat tcc cag agt gct gat tgg 768
Ala Pro Glu Val Val Asn Arg Arg Gly His Ser Gln Ser Ala Asp Trp
245 250 255

tgg tca tat ggt gtt ctt atg ttt gaa atg ctt act ggt act ctg cca 816
Trp Ser Tyr Gly Val Leu Met Phe Glu Met Leu Thr Gly Thr Leu Pro
260 265 270

ttt caa ggt aaa gac aga aat gag acc atg aat atg ata tta aaa gca 864
Phe Gln Gly Lys Asp Arg Asn Glu Thr Met Asn Met Ile Leu Lys Ala
275 280 285

aaa ctt gga atg cct caa ttt ctt agt gct gaa gca caa agt ctt cta 912
Lys Leu Gly Met Pro Gln Phe Leu Ser Ala Glu Ala Gln Ser Leu Leu
290 295 300

agg atg tta ttc aaa agg aat cca gca aat aga ttg gga tca gaa gga 960
Arg Met Leu Phe Lys Arg Asn Pro Ala Asn Arg Leu Gly Ser Glu Gly
305 310 315 320

gtt gaa gaa atc aaa aga cat ctg ttt ttt gca aat att gac tgg gat 1008

Val Glu Glu Ile Lys Arg His Leu Phe Phe Ala Asn Ile Asp Trp Asp
325 330 335

aaa tta tat aaa aga gaa gtt caa cct cct ttc aaa cct gct tct gga 1056
Lys Leu Tyr Lys Arg Glu Val Gln Pro Pro Phe Lys Pro Ala Ser Gly
340 345 350

aaa cca gat gat act ttt tgt ttt gat cct gaa ttt act gca aaa aca 1104
Lys Pro Asp Asp Thr Phe Cys Phe Asp Pro Glu Phe Thr Ala Lys Thr
355 360 365

cct aaa gat tct ccc ggt ttg cca gcc agt gca aat gct cat cag ctc 1152
Pro Lys Asp Ser Pro Gly Leu Pro Ala Ser Ala Asn Ala His Gln Leu
370 375 380

ttc aaa gga ttc agc ttt gtt gca act tct att gca gaa gaa tat aaa 1200
Phe Lys Gly Phe Ser Phe Val Ala Thr Ser Ile Ala Glu Glu Tyr Lys
385 390 395 400

atc act cct atc aca agt gca aat gta tta cca att gtt cag ata aat 1248
Ile Thr Pro Ile Thr Ser Ala Asn Val Leu Pro Ile Val Gln Ile Asn
405 410 415

gga aat gct gca caa ttt ggt gaa gta tat gaa ttg aag gag gat att 1296
Gly Asn Ala Ala Gln Phe Gly Glu Val Tyr Glu Leu Lys Glu Asp Ile
420 425 430

ggt gtt ggc tcc tac tct gtt tgc aag cga tgc ata cat gca act acc 1344
Gly Val Gly Ser Tyr Ser Val Cys Lys Arg Cys Ile His Ala Thr Thr
435 440 445

aac atg gaa ttt gca gtg aag atc att gac aaa agt aag cga gac cct 1392
Asn Met Glu Phe Ala Val Lys Ile Ile Asp Lys Ser Lys Arg Asp Pro
450 455 460

tca gaa gag att gaa ata ttg atg cgc tat gga caa cat ccc aac att 1440
Ser Glu Glu Ile Glu Ile Leu Met Arg Tyr Gly Gln His Pro Asn Ile
465 470 475 480

att act ttg aag gat gtc ttt gat gat ggt aga tat gtt tac ctt gtt 1488
Ile Thr Leu Lys Asp Val Phe Asp Asp Gly Arg Tyr Val Tyr Leu Val
485 490 495

acg gat tta atg aaa gga gga gag tta ctt gac cgt att ctc aaa caa 1536
Thr Asp Leu Met Lys Gly Gly Glu Leu Leu Asp Arg Ile Leu Lys Gln
500 505 510

aaa tgt ttc tcg gaa cgg gag gct agt gat ata cta tat gta ata agt 1584
Lys Cys Phe Ser Glu Arg Glu Ala Ser Asp Ile Leu Tyr Val Ile Ser
515 520 525

aag aca gtt gac tat ctt cat tgt caa gga gtt gtt cat cgt gat ctt 1632
Lys Thr Val Asp Tyr Leu His Cys Gln Gly Val Val His Arg Asp Leu
530 535 540

aaa cct agt aat att tta tac atg gat gaa tca gcc agt gca gat tca 1680
Lys Pro Ser Asn Ile Leu Tyr Met Asp Glu Ser Ala Ser Ala Asp Ser
545 550 555 560

atc agg ata tgt gat ttt ggg ttt gca aaa caa ctt cga gga gaa aat 1728
Ile Arg Ile Cys Asp Phe Gly Phe Ala Lys Gln Leu Arg Gly Glu Asn

565 570 575

gga ctt ctc tta act cca tgc tac act gca aac ttt gtt gca cct gag 1776
Gly Leu Leu Leu Thr Pro Cys Tyr Thr Ala Asn Phe Val Ala Pro Glu
580 585 590

gtt ctt atg caa cag gga tat gat gct gct tgt gat atc tgg agt tta 1824
Val Leu Met Gln Gln Gly Tyr Asp Ala Ala Cys Asp Ile Trp Ser Leu
595 600 605

gga gtc ctt ttt tac aca atg ttg gct ggc tac act cca ttt gct aat 1872
Gly Val Leu Phe Tyr Thr Met Leu Ala Gly Tyr Thr Pro Phe Ala Asn
610 615 620

ggc ccc aat gat act cct gaa gag ata ctg ctg cgt ata ggc aat gga 1920
Gly Pro Asn Asp Thr Pro Glu Glu Ile Leu Leu Arg Ile Gly Asn Gly
625 630 635 640

aaa ttc tct ttg agt ggt gga aac tgg gac aat att tca gac gga gca 1968
Lys Phe Ser Leu Ser Gly Gly Asn Trp Asp Asn Ile Ser Asp Gly Ala
645 650 655

aag gat ttg ctt tcc cat atg ctt cat atg gac cca cat cag cgg tat 2016
Lys Asp Leu Leu Ser His Met Leu His Met Asp Pro His Gln Arg Tyr
660 665 670

act gct gaa caa ata tta aag cac tca tgg ata act cac aga gac cag 2064
Thr Ala Glu Gln Ile Leu Lys His Ser Trp Ile Thr His Arg Asp Gln
675 680 685

ttg cca aat gat cag cca aag aga aat gat gtg tca cat gtt gtt aag 2112
Leu Pro Asn Asp Gln Pro Lys Arg Asn Asp Val Ser His Val Val Lys
690 695 700

gga gca atg gtt gca aca tac tct gcc ctg act cac aag acc ttt caa 2160
Gly Ala Met Val Ala Thr Tyr Ser Ala Leu Thr His Lys Thr Phe Gln
705 710 715 720

cca gtc cta gag cct gta gct gct tca agc tta gcc cag cga cgg agc 2208
Pro Val Leu Glu Pro Val Ala Ala Ser Ser Leu Ala Gln Arg Arg Ser
725 730 735

atg aaa aag cga aca tca act ggc ctg 2235
Met Lys Lys Arg Thr Ser Thr Gly Leu
740 745

<210> 4
<211> 745
<212> PRT
<213> Homo sapiens

<400> 4

Met Leu Pro Phe Ala Pro Gln Asp Glu Pro Trp Asp Arg Glu Met Glu
1 5 10 15

Val Phe Ser Gly Gly Gly Ala Ser Ser Gly Glu Val Asn Gly Leu Lys
20 25 30

Met Val Asp Glu Pro Met Glu Glu Gly Glu Ala Asp Ser Cys His Asp
35 40 45

Glu Gly Val Val Lys Glu Ile Pro Ile Thr His His Val Lys Glu Gly
50 55 60

Tyr Glu Lys Ala Asp Pro Ala Gln Phe Glu Leu Leu Lys Val Leu Gly
65 70 75 80

Gln Gly Ser Phe Gly Lys Val Phe Leu Val Arg Lys Lys Thr Gly Pro
85 90 95

Asp Ala Gly Gln Leu Tyr Ala Met Lys Val Leu Lys Lys Ala Ser Leu
100 105 110

Lys Val Arg Asp Arg Val Arg Thr Lys Met Glu Arg Asp Ile Leu Val
115 120 125

Glu Val Asn His Pro Phe Ile Val Lys Leu His Tyr Ala Phe Gln Thr
130 135 140

Glu Gly Lys Leu Tyr Leu Ile Leu Asp Phe Leu Arg Gly Gly Asp Val
145 150 155 160

Phe Thr Arg Leu Ser Lys Glu Val Leu Phe Thr Glu Glu Asp Val Lys
165 170 175

Phe Tyr Leu Ala Glu Leu Ala Leu Ala Leu Asp His Leu His Gln Leu
180 185 190

Gly Ile Val Tyr Arg Asp Leu Lys Pro Glu Asn Ile Leu Leu Asp Glu
195 200 205

Ile Gly His Ile Lys Leu Thr Asp Phe Gly Leu Ser Lys Glu Ser Val
210 215 220

Asp Gln Glu Lys Lys Ala Tyr Ser Phe Cys Gly Thr Val Glu Tyr Met
225 230 235 240

Ala Pro Glu Val Val Asn Arg Arg Gly His Ser Gln Ser Ala Asp Trp
245 250 255

Trp Ser Tyr Gly Val Leu Met Phe Glu Met Leu Thr Gly Thr Leu Pro
260 265 270

```
Phe Gln Gly Lys Asp Arg Asn Glu Thr Met Asn Met Ile Leu Lys Ala
        275                 280                 285

Lys Leu Gly Met Pro Gln Phe Leu Ser Ala Glu Ala Gln Ser Leu Leu
    290                 295                 300

Arg Met Leu Phe Lys Arg Asn Pro Ala Asn Arg Leu Gly Ser Glu Gly
305                 310                 315                 320

Val Glu Glu Ile Lys Arg His Leu Phe Phe Ala Asn Ile Asp Trp Asp
                325                 330                 335

Lys Leu Tyr Lys Arg Glu Val Gln Pro Pro Phe Lys Pro Ala Ser Gly
            340                 345                 350

Lys Pro Asp Asp Thr Phe Cys Phe Asp Pro Glu Phe Thr Ala Lys Thr
        355                 360                 365

Pro Lys Asp Ser Pro Gly Leu Pro Ala Ser Ala Asn Ala His Gln Leu
    370                 375                 380

Phe Lys Gly Phe Ser Phe Val Ala Thr Ser Ile Ala Glu Glu Tyr Lys
385                 390                 395                 400

Ile Thr Pro Ile Thr Ser Ala Asn Val Leu Pro Ile Val Gln Ile Asn
                405                 410                 415

Gly Asn Ala Ala Gln Phe Gly Glu Val Tyr Glu Leu Lys Glu Asp Ile
            420                 425                 430

Gly Val Gly Ser Tyr Ser Val Cys Lys Arg Cys Ile His Ala Thr Thr
        435                 440                 445

Asn Met Glu Phe Ala Val Lys Ile Ile Asp Lys Ser Lys Arg Asp Pro
    450                 455                 460

Ser Glu Glu Ile Glu Ile Leu Met Arg Tyr Gly Gln His Pro Asn Ile
465                 470                 475                 480

Ile Thr Leu Lys Asp Val Phe Asp Asp Gly Arg Tyr Val Tyr Leu Val
                485                 490                 495

Thr Asp Leu Met Lys Gly Gly Glu Leu Leu Asp Arg Ile Leu Lys Gln
            500                 505                 510

Lys Cys Phe Ser Glu Arg Glu Ala Ser Asp Ile Leu Tyr Val Ile Ser
```

```
            515                 520                 525

Lys Thr Val Asp Tyr Leu His Cys Gln Gly Val Val His Arg Asp Leu
    530                 535                 540

Lys Pro Ser Asn Ile Leu Tyr Met Asp Glu Ser Ala Ser Ala Asp Ser
545                 550                 555                 560

Ile Arg Ile Cys Asp Phe Gly Phe Ala Lys Gln Leu Arg Gly Glu Asn
                565                 570                 575

Gly Leu Leu Leu Thr Pro Cys Tyr Thr Ala Asn Phe Val Ala Pro Glu
            580                 585                 590

Val Leu Met Gln Gln Gly Tyr Asp Ala Ala Cys Asp Ile Trp Ser Leu
        595                 600                 605

Gly Val Leu Phe Tyr Thr Met Leu Ala Gly Tyr Thr Pro Phe Ala Asn
    610                 615                 620

Gly Pro Asn Asp Thr Pro Glu Glu Ile Leu Leu Arg Ile Gly Asn Gly
625                 630                 635                 640

Lys Phe Ser Leu Ser Gly Gly Asn Trp Asp Asn Ile Ser Asp Gly Ala
                645                 650                 655

Lys Asp Leu Leu Ser His Met Leu His Met Asp Pro His Gln Arg Tyr
            660                 665                 670

Thr Ala Glu Gln Ile Leu Lys His Ser Trp Ile Thr His Arg Asp Gln
        675                 680                 685

Leu Pro Asn Asp Gln Pro Lys Arg Asn Asp Val Ser His Val Val Lys
    690                 695                 700

Gly Ala Met Val Ala Thr Tyr Ser Ala Leu Thr His Lys Thr Phe Gln
705                 710                 715                 720

Pro Val Leu Glu Pro Val Ala Ala Ser Ser Leu Ala Gln Arg Arg Ser
                725                 730                 735

Met Lys Lys Arg Thr Ser Thr Gly Leu
            740                 745

<210>  5
<211>  22
```

<212> DNA
<213> Homo sapiens

<400> 5
ggagcaaagg gagcaatggt tg 22

<210> 6
<211> 22
<212> DNA
<213> Homo sapiens

<400> 6
tcttcatcca gtttggccta gg 22

<210> 7
<211> 20
<212> DNA
<213> Homo sapiens

<400> 7
tgggtgtgaa ccatgagaag 20

<210> 8
<211> 20
<212> DNA
<213> Homo sapiens

<400> 8
gtgtcgctgt tgaagtcaga 20